Ecological Informatics 17 (2013) 30–35

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Habitat suitability modelling for mayflies (Ephemeroptera) in Flanders (Belgium)

Koen Lock *, Peter L.M. Goethals

Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, Belgium

ARTICLE INFO

Article history:
Received 18 February 2011
Received in revised form 12 July 2011
Accepted 29 December 2011
Available online 5 January 2012

Keywords:
Artificial neural networks
Classification trees
Logistic regressions
Random forests
Support vector machines

ABSTRACT

Logistic regressions, artificial neural networks, support vector machines, random forests and classification trees were used to predict habitat suitability for mayflies on basis of physical–chemical water characteristics, dominant land use type and structural parameters. As a case-study, the dataset of Flanders was used, which consisted of 4289 samples containing mayflies (presences) and 3315 samples from waters where mayflies were never encountered (absences). For all techniques, data were randomly divided in a training set (two thirds) and a test set (one third). Models were calibrated using a tenfold cross-validation on the training set and subsequently validated using the test set. All techniques delivered good models that were able to discriminate sites with and without mayflies and performance (expressed as percent correctly classified instances and kappa-statistics) was in all cases similar for the training and the test set. Artificial neural networks and random forests performed slightly better compared to the other techniques. Samples with mayflies contained significantly more oxygen, a better developed river structure, higher values for sinuosity and steeper slopes, while samples without mayflies had significantly higher values for ammonium, nitrite, Kjeldahl nitrogen, total phosphorous, orthophosphate, biological and chemical oxygen demand, pH and conductivity. Also land use differed significantly, with mayflies usually present in forests but absent in industrial areas. The prevalence of mayflies gradually increased during the nineties from about 20 to 40%, which corresponded with an improvement of the chemical water quality. During the last decade, however, water quality did not further improve and as a result, mayflies prevalence did not continue to increase. Based on the planned measures, an ensemble forecast using the five mentioned modelling techniques predicted that mayflies prevalence will increase to 46% by 2015 and to 72% by 2027. To meet the requirements of the European Water Framework Directive, which states that all surface waters should obtain a good ecological quality, extra efforts will be needed to decrease nutrient concentrations and to improve habitat quality.

1. Introduction

To achieve a good ecological status of groundwater and surface water in Europe by 2015 is the main objective of the European Union Water Framework Directive (WFD; European Council, 2000). The WFD requires the use of biotic indicators such as macrobenthic fauna, fish fauna and aquatic flora to assess the ecological water quality. All member-states have to develop methods that are concurrent with the WFD. The presented case-study was performed in Flanders, where the Multimetric Macroinvertebrate Index Flanders (MMIF; Gabriels et al., 2010) was recently developed in order to meet the requirements of the WFD. In this index, mayflies are recognised as a sensitive group of water invertebrates, which occur in waters with a good ecological water quality. Despite their importance as water quality indicators, mayflies only recently received attention in Flanders (Lock & Goethals, 2011).

In Belgium, nature conservation policy is the responsibility of the regional governments (Flanders, Brussels and Wallonia). In the presented study, we focussed on the situation in Flanders, the Dutch speaking northern part of Belgium, where the water quality is monitored by the Flemish Environment Agency. As in most European countries, river management in Flanders has been conducted at the river basin level using wastewater treatment plants and imposed standards of effluent concentrations. Although these measures resulted in a significant improvement of the chemical and ecological water quality since the beginning of the nineties (VMM, 2010), many Flemish surface waters still lack the required good ecological status. This is also reflected by the limited distribution of most mayfly species in Flanders (Lock & Goethals, 2011). In order to efficiently allocate restoration efforts, ecological models could be useful for river managers and stakeholders to select among different restoration options and management strategies (Mouton et al., 2009).

Because one modelling technique that is efficient for one dataset might fail for another, it is not possible to identify *the* best modelling technique. Therefore, Thuiller et al. (2009) proposed that ensemble forecasts with several modelling techniques should be made and that the resulting range of predictions should be analysed rather

* Corresponding author at: Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, J. Plateaustraat 22, 9000 Ghent, Belgium. Tel.: +32 9 2643996; fax: +32 9 2643766.
E-mail address: Koen_Lock@hotmail.com (K. Lock).

1574-9541/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoinf.2011.12.004

than relying on the results of a single model. In the present study, habitat suitability models were used to predict the presence or absence of mayflies in surface waters with five modelling techniques: logistic regressions, artificial neural networks, support vector machines, random forests and classification trees. With the exception of support vector machines, which only recently became popular for modelling ecological data (Ambelu et al., 2010; Hoang et al., 2010; Pino-Mejias et al., 2010), all these modelling techniques have been frequently used for habitat suitability modelling. Not only the traditional physical–chemical parameters were used to model the presence of mayflies, but it was also investigated whether other parameters such as land use type and river morphology could improve model performance. Subsequently, these five modelling techniques were used to make an ensemble forecast of the mayfly prevalence in 2015 and 2027 based on modelled future oxygen and nutrient concentrations.

2. Materials and methods

2.1. Macroinvertebrate sampling

Since 1989, macroinvertebrates have been sampled at several thousand sampling locations in Flanders (northern part of Belgium) in the context of water quality monitoring by the Flemish Environment Agency. During monitoring, macroinvertebrates were sampled by kicksampling using a standard handnet, as described by Gabriels et al. (2010). During a previous study, mayflies were identified to species level (Lock & Goethals, 2011). As most species were very rare in Flanders, it was unfeasible to develop species-specific habitat suitability models for all species separately. Hence, it was decided to model the presence or absence of mayflies in general in the present study. All types of stagnant and running waters in Flanders could potentially contain mayflies and their current absence in most waters is only caused by human disturbances.

2.2. Environmental parameters

Dissolved oxygen, conductivity and pH were always measured in the field during macroinvertebrate sampling. The other chemical parameters (Table 1) were retrieved from monitoring data of the chemical water quality of the Flemish surface waters, which is also performed by the Flemish Environment Agency. Because the chemical monitoring, which was usually performed on a monthly basis, was not performed simultaneously with the macroinvertebrate sampling, measurements from the last date before macroinvertebrate sampling were used. The slope of a watercourse was determined based on the difference in height between two points 1000 m apart using GIS-software applied on the Flemish Hydrographic Atlas (AGIV, 2006). The same data were used to determine the sinuosity on a stretch of 100 m. River morphology was evaluated based on pictures of the sampling sites: pool-riffle pattern and meandering were both quoted from 0 (absent) to 5 (well developed) and summed, which yielded a score from 0 to 10. Aerial photographs provided by Google Earth were used to distinguish five different land use types. Along a stretch of 500 m upstream of the sampling point, the main land use type was assigned to forest, meadow, arable land, urban or industry. Differences in environmental parameters between sites where mayflies were present and absent were analysed using Kruskal–Wallis ANOVA (StatSoft, 2004). Chi-square tests were used to test the difference in land use types between sites with and without mayflies.

Table 1
Median values (with 10 and 90 percentiles) of the assessed environmental parameters in samples where mayflies were present or absent (BOD: Biological Oxygen Demand; COD: Chemical Oxygen Demand).

		Present	Absent
pH		7.66 (6.98–8.17)	7.7 (7.21–8.26)
Conductivity	μS/cm	578 (304–1010)	878 (444–2990)
Oxygen	mg/L	7.4 (3.8–10)	5.6 (2.2–10)
BOD	mg/L	3.0 (2.0–7.0)	5.0 (2.5–12)
COD	mg/L	25 (11–58)	38 (19–82)
Ammonium	mg/L	0.50 (0.11–2.9)	1.8 (0.25–8.2)
Nitrite	mg/L	0.088 (0.020–0.31)	0.17 (0.029–0.53)
Nitrate	mg/L	2.6 (0.48–7.3)	2.8 (0.23–8.8)
Kjeldal nitrogen	mg/L	2.3 (1.1–5.1)	4.0 (1.6–11)
Phosphorus	mg/L	0.50 (0.16–1.2)	0.91 (0.31–2.3)
Orthophosphate	mg/L	0.17 (0.06–0.67)	0.46 (0.10–1.6)
Slope	m/1000 m	0.91 (0.11–5.6)	0.71 (0.074–3.5)
Sinuosity		1.02 (1.00–1.19)	1.01 (1.00–1.15)
River structure	Classes from 0–10	3 (2–5)	2 (0–4)

2.3. Modelling

In total, mayflies were present in 4289 samples from Flemish surface waters and as absence data, 3315 samples were considered from sites where mayflies have never been observed. The dataset was randomly split in two thirds for training and one third for validation. During calibration, a tenfold cross-validation was performed to avoid overfitting. Five modelling techniques were applied to model the presence/absence of mayflies: logistic regressions, classification trees, random forests, artificial neural networks and support vector machines. Logistic regressions (LR) are used for the prediction of the probability of occurrence of an event by fitting data to a logit function logistic curve. Here, a multinomial logistic regression with ridge estimator was used as a generalised linear model. Classification trees (CT) summarise the relationships between explanatory variables and the response variable (presence/absence of mayflies) in a dichotomously branching tree. Each bifurcation is defined by a certain value of one of the explanatory variables dividing the dataset in two more homogenous subsets. CT were grown automatically using the J48 algorithm that minimises the impurity of the subsets (Witten and Frank, 2005). A fully grown tree would explain the training data with a high accuracy, but it would fail for unseen data because of overfitting. CT generality was increased by pruning, which yielded simpler trees that usually result in better classification of unseen data (Dakou et al., 2007). Random forests (RF) are an ensemble learning alternative to CT: many trees are constructed, with classes that are predicted by majority vote. A RF is grown by a procedure called bagging, which is short for bootstrap aggregating, where each tree is independently constructed by using a bootstrap sample (with replacement) of the entire dataset. Each node of the trees is split using only a subset of the explanatory variables chosen randomly for each tree. Artificial Neural Networks (ANN) are non-linear statistical data modelling tools that are based on the architecture of biological neural networks and consist of a group of interconnected computing units or neurons. During a learning phase, connection weights among the neurons are adapted by backpropagating training data through the net. Support Vector Machines (SVM) are developed from a linear classifier using a maximum hyperplane to separate two classes. In a non-linear case, the central idea of classification with SVM is to map training data into a higher-dimensional feature space and to compute separating hyperplanes that achieve maximum separation between classes. The maximum separation hyperplane is only a function of the training data that lie on the margin and are called support vectors. Platt's sequential minimal optimisation algorithm (Keerthi et al., 2001) was used for training a support-vector classifier, which replaces all missing values, transforms nominal attributes into binary ones and multi-class problems are solved using pairwise classification. All modelling techniques were performed using WEKA software (Witten and Frank, 2005). Four sets of models were developed, which were either based on physical–chemical parameters, land use types, structural characteristics of the surface waters or all these parameters together. To evaluate the performance of each

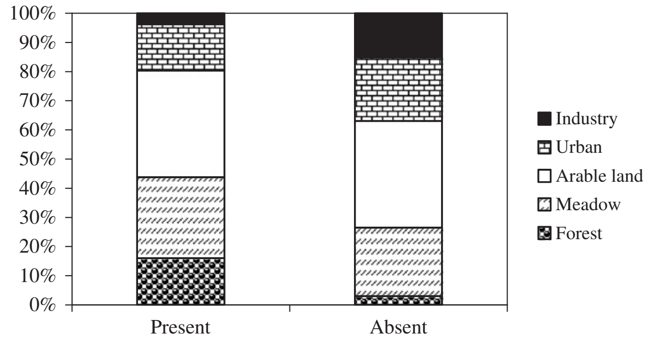

Fig. 1. Land use at the sampling locations where mayflies were present and absent.

Table 2
Correctly Classified Instances (CCI) and Cohen's kappa statistics (*K*) for calibration and validation with five different modelling techniques that were used to predict the presence/absence of mayflies based on physical–chemical parameters only and based on all measured variables.

	Calibration		Validation	
	CCI (%)	K	CCI (%)	K
Physico-chemistry				
Logistic regressions	69	0.37	70	0.37
Artificial neural networks	72	0.44	74	0.47
Support vector machines	69	0.36	69	0.37
Random forests	67	0.33	68	0.35
Classification trees	69	0.36	70	0.36
All variables				
Logistic regressions	70	0.39	72	0.42
Artificial neural networks	73	0.45	75	0.49
Support vector machines	70	0.38	71	0.40
Random forests	73	0.47	74	0.48
Classification trees	68	0.34	70	0.37

technique, the percentage of correctly classified instances (% CCI) and Cohen's kappa statistics (*K*) were used (Witten and Frank, 2005).

2.4. Ensemble forecast

Based on the planned measures, especially collecting and treating a higher fraction of the domestic waste water, oxygen and nutrient concentrations were modelled using the water quality model PEGASE (VMM, 2009b). More specifically, models were developed using the five mentioned techniques based on the concentration of oxygen, biological oxygen demand, nitrate, ammonium, Kjeldal nitrogen, phosphate and orthophosphate. An ensemble forecast based on the five techniques was used to model mayfly prevalence in a reference situation in 2006 and two future scenarios in 2015 and 2027.

3. Results

3.1. Environmental parameters

Kruskal–Wallis ANOVA indicated that pH, conductivity, biological oxygen demand, chemical oxygen demand, ammonium, nitrite, Kjeldahl nitrogen, phosphorus, orthophosphate were significantly lower (all $p<0.001$) and dissolved oxygen, slope, river morphology and sinuosity were significantly higher (all $p<0.001$) in samples with mayflies compared to samples without mayflies. No significant differences in nitrate were observed between samples with and without mayflies ($p=0.67$). Median values as well as 10 and 90 percentiles of all environmental parameters are listed in Table 1. The land use type differed significantly (Chi-Square, $p<0.001$) between samples with and without mayflies. While mayflies were usually present in forests, they were rarely present in industrial areas (Fig. 1).

3.2. Modelling

Presence and absence of mayflies could be quite accurately modelled based on physical–chemical parameters with all five modelling techniques (Table 2). However, models based on the type of land use or based on structural parameters (slope, sinuosity and river morphology) did not perform well, with hardly 60% correctly classified instances and Cohen's kappa lower than 0.25 for all models. When all variables were used simultaneously, the accuracy of the models was only slightly higher in comparison with models based on physical–chemical parameters alone, with in most cases more than 70% correctly classified instances and Cohen's kappa higher than 0.4 (Table 2). Artificial neural networks and random forests gave the best predictions, while classification trees performed somewhat less accurate. An example of a strongly pruned classification tree is presented in Fig. 2.

3.3. Temporal variation

The prevalence of mayflies gradually increased during the nineties from about 20 to almost 40% (Fig. 3). This increase corresponded with an improvement of the chemical water quality as indicated by for example decreased orthophosphate and ammonium concentrations (Fig. 3). During the last decade, however, water quality did not

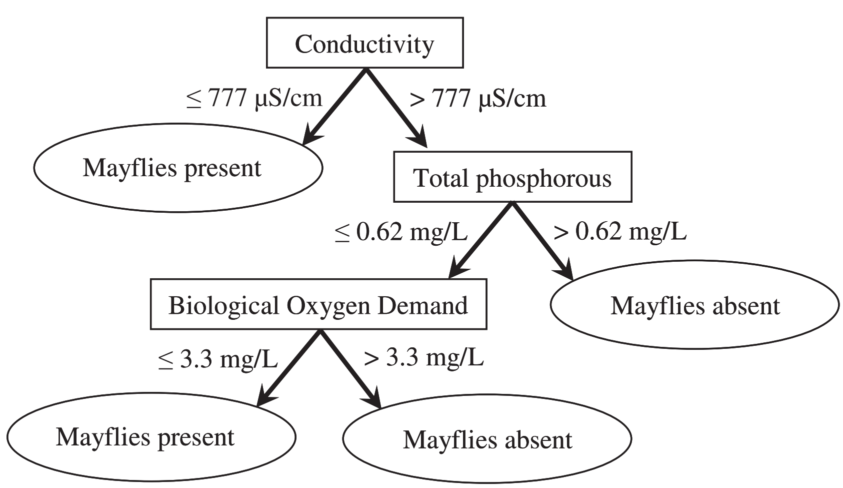

Fig. 2. Strongly pruned classification tree that predicts the presence or absence of mayflies in Flemish surface waters (correctly predicted instances 67%, Cohen's kappa 0.32).

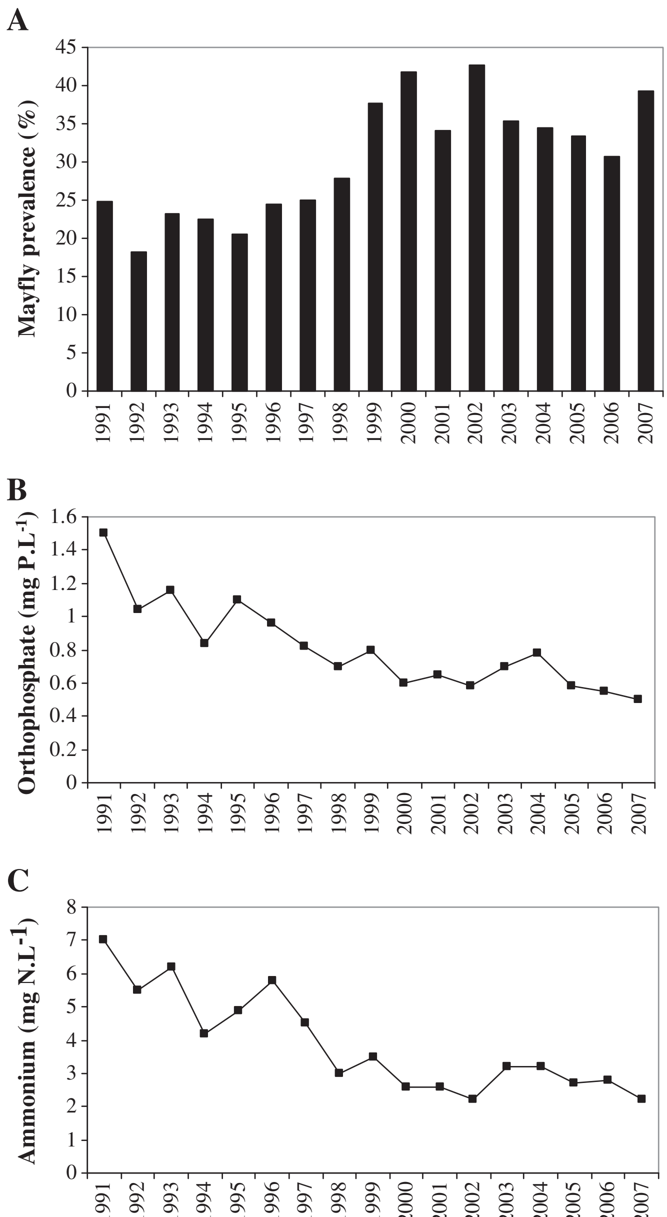

Fig. 3. The prevalence of the mayflies (A), the average orthophosphorus concentration (B) and average the ammonium concentration (C) in Flemish surface waters from 1991 till 2007.

further improve and as a result, mayfly prevalence did not continue to increase (Fig. 3).

3.4. Ensemble forecast

Based on modelled oxygen and nutrient concentrations, the five mentioned modelling techniques were used to make an ensemble forecast of the mayfly prevalence. The modelled mayfly prevalence increased from less than 40% in the reference year 2006 to 46% in 2015 and 72% in 2027 (Fig. 4).

4. Discussion

The present case-study was performed in Flanders, which has a population of 6.2 million inhabitants and a high population density of 456 citizens/km². About 88% is connected to a sewage system, however, only 70.3% is actually treated (VMM, 2009a). As rainwater

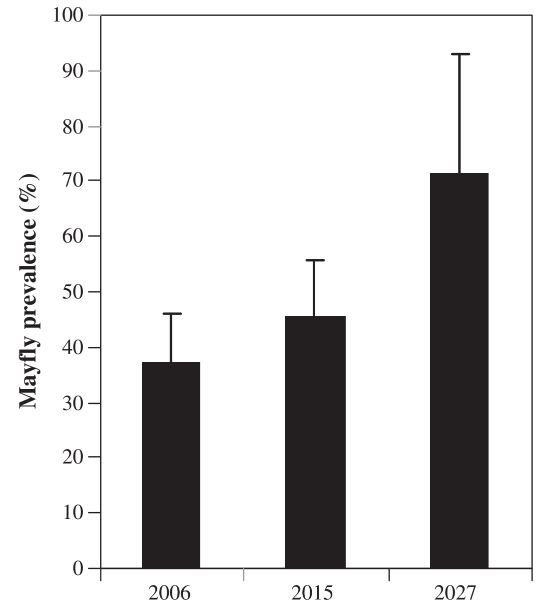

Fig. 4. Ensemble forecast of mayfly prevalence (with standard deviation) in 2006, 2015 and 2027 based on water characteristics modelled with PEGASE.

is often not collected separately, untreated water is regularly discharged after heavy rains, which results in problematic drops in dissolved oxygen concentration and peak levels of substances such as ammonium. Flanders is also heavily industrialised and the intensive agriculture causes a heavy pressure as 53% of the land is used for agriculture (VMM, 2009a). In addition, thousands of weirs have been built for flood control, hundreds of kilometres of artificial banks have been installed and the majority of the river channels has been straightened. Although the situation in Flanders is probably more problematic than in most other countries in Europe due to the high population density, the other member-states have to deal with similar problems.

In Flanders, most attention is currently focused on the restoration of watercourses with the poorest water quality, however, ameliorating water quality from bad to poor or moderate will not help populations of sensitive organisms such as most mayfly species. Recently, most surface waters that had a bad or poor ecological water quality for macroinvertebrates still improved, however, more than half of the locations that had a very good score for macroinvertebrates declined (VMM, 2009a). A first step should therefore be to protect the sites which still have a high ecological water quality and contain a lot of sensitive organisms such as mayflies. As most suitable habitats are now isolated, populations are extremely vulnerable to extinction and recolonisation is hardly possible (Lock & Goethals, 2011). Therefore, also intentional interventions are needed in order to connect isolated populations by solving the present bottlenecks that prevent the expansion of remaining populations. To efficiently allocate restoration efforts, ecological models are useful for the assessment of these bottlenecks in the river basin (Mouton et al., 2008).

During the nineties, the prevalence of mayflies gradually increased, which could be related to the improved water quality (Fig. 3). This increased prevalence was especially due to the mayfly species *Baetis vernus*. *B. vernus* is known to be a relatively tolerant species (Beketov, 2004; Tixier et al., 2009) and it is the only common species of running waters in Flanders (Lock & Goethals, 2011). The other mayfly species occurring in running waters did not significantly increase during the last twenty years, which is probably caused by their higher sensitivity. A further improvement of the water quality is needed for the latter species in order to allow an increase of their prevalence.

In the present study, all five modelling techniques resulted in reliable models, only the performance of classification trees was slightly less accurate (Table 2). In Fig. 2, an example of a strongly pruned classification tree is presented, which indicates that mayflies are always present in waters with a low conductivity, but if the conductivity is high, they are only present when both the phosphate concentration and the biological oxygen demand are low. This example illustrates that classification trees can be easily understood and communicated. Despite its slightly lower accuracy, this technique is very suitable to convince river managers, decision makers or even the public (Boets et al., 2010; Dominguez-Granda et al., 2011).

The occurrence of mayflies could be predicted based on the physical–chemical water characteristics and further improved when land-use and variables describing river morphology were taken into account. Land-use appears to be a key factor influencing macroinvertebrate community composition among sites (Collier, 2008; Compin and Cereghino, 2007; Sponseller et al., 2001; Strayer et al., 2003). Urban and industrial sites in the river basin represented the pressure with the most negative impact on macroinvertebrate indices, while forests and pastures had a positive effect (Wasson et al., 2010). An advantage of the parameters describing land use and river morphology is that they are quite stable in time. On the other hand, physical–chemical parameters can strongly fluctuate, which can easily lead to false conclusions regarding habitat suitability. Including land use and river morphology can therefore improve the accuracy of habitat suitability models. However, in the present study, the latter variables only slightly improved model performance. This was probably caused by the fact that a few mayfly species are relatively tolerant and occur already in waters with a moderate quality.

The developed models could distinguish suitable from unsuitable habitats for mayflies based on physical–chemical variables, whereas including land use and morphological characteristics only slightly improved model performance. Although some mayfly species already occur in waters with a moderate quality, most surface waters in Flanders are still not good enough to allow their occurrence, since their prevalence is still hardly 40%. An ensemble forecast with the five modelling techniques indicated that mayfly prevalence will increase from 37% in 2006 to 46% in 2015 and 72% in 2027 when the planned measures are carried out (Fig. 4), which clearly indicates a lack of ambition. The scenario's of 2015 and 2027 are mainly based on the planned installation of waste water treatment plants, however, since the Flemish government tends to plan more than is actually carried out in the field, it can be expected that the water quality will not even improve as much as predicted if no additional measures are taken. The modelled prevalence in 2006 closely corresponds with the values observed during the last decade (Fig. 3), which reflects that the ensemble forecast gave a good estimate of the mayfly prevalence. However, in the same reference year 2006, less than 10% of the surface waters reached a good ecological water quality for macroinvertebrates (VMM, 2009b), indicating that the presence of mayflies is not sufficient to obtain a good ecological water quality. In order to obtain a good water quality in all Flemish surface waters, which should be the case by 2015 (or at the very latest by 2027 if this is not feasible) according to the WFD (European Council, 2000), there is thus still a lot of work to be done.

De Cooman et al. (2007) indicated that the goals of the WFD could only be achieved by implementing small-scale efforts such as natural bank restoration, fish passage construction or river channel re-meandering, which affect physical and chemical habitats both locally and at the basin scale. Although this kind of measures are undoubtedly beneficial, they are also very expensive. Especially re-meandering is costly and our data even indicated that sinuosity has only a minor effect on the occurrence of mayflies. Jahnig et al. (2010) concluded that habitat restoration within a small stretch is generally not sufficient to realize changes in benthic invertebrate community composition and that restoring habitat on a larger scale, using more comprehensive measures and tackling catchment-wide problems (e.g. water quality, source populations) are required for a recovery of the invertebrate community. More cost-effective measures might therefore be the installation of constructed wetlands for small-scale waste water treatment and the creation of buffer zones. Currently, agricultural land in Flanders usually extends up to the river banks and a buffer zones are rarely present although these are known to decrease the runoff of nutrients and pesticides (Sahu and Gu, 2009; Tran et al., 2010). Wasson et al. (2010) indicated that riparian corridors are manageable areas and their creation along European watercourses should receive priority in order to achieve a good ecological status. On the other hand, constructed wetlands are considered as a cost-effective alternative for the treatment of point-sources that can not easily be connected to a sewage treatment plant (Meers et al., 2008; Boets et al., 2011).

5. Conclusion

Although artificial neural networks and random forests performed slightly better, all used modelling techniques were able to distinguish sites where mayflies were present or absent based on physical–chemical parameters. Adding land-use and structural parameters hardly improved model performance. Mayfly prevalence increased from 20 to 40% during the nineties, however, no further improvement was detected during the last decade. The modelled future improvement of the water quality expected on the basis of the planned waste water treatment plants was used to predict mayfly prevalence using a ensemble forecasting strategy based on the five used modelling techniques. It was predicted that mayfly prevalence will further increase, however, the water quality improvement will probably not be sufficient to obtain the good ecological quality in all Flemish surface waters as required by the WFD (European Council, 2000). Therefore, it is suggested to apply additional measures besides the installation of waste water treatment plants, such as the creation of buffer zones and the installation of constructed wetlands.

Acknowledgments

We would like to thank the Flemish Environment Agency (VMM) for their environmental data and for the opportunity to study their samples. Koen Lock is currently supported by a post-doctoral fellowship from the Fund for Scientific Research (FWO-Vlaanderen, Belgium).

References

AGIV, 2006. Vlaamse Hydrografische Atlas. Agentschap voor Geografische Informatie Vlaanderen, Gent. http://geo-vlaanderen.agiv.be/geo-vlaanderen/vha.

Ambelu, A., Lock, K., Goethals, P.L.M., 2010. Comparison of modeling techniques to predict macroinvertebrate community composition in rivers of Ethiopia. Ecological Informatics 5, 147–152.

Beketov, M.A., 2004. Different sensitivity of mayflies (Insecta, Ephemeroptera) to ammonia, nitrite and nitrate: linkage between experimental and observational data. Hydrobiologia 528, 209–216.

Boets, P., Lock, K., Messiaen, M., Goethals, P.L.M., 2010. Combining datadriven methods and lab studies to analyse the ecology of *Dikerogammarus villosus*. Ecological Informatics 5, 133–139.

Boets, P., Michels, E., Meers, E., Lock, K., Tack, F.M.G., Goethals, P.L.M., 2011. Integration of landscaping and ecological development in constructed wetlands for manure treatment. Wetlands 31, 763–771.

Collier, K.J., 2008. Temporal patterns in the stability, persistence and condition of stream macroinvertebrate communities: relationships with catchment land-use and regional climate. Freshwater Biology 53, 603–616.

Compin, A., Cereghino, R., 2007. Spatial patterns of macroinvertebrate functional feeding groups in streams in relation to physical variables and land-cover in Southwestern France. Landscape Ecology 22, 1215–1225.

Dakou, E., D'heygere, T., Dedecker, A.P., Goethals, P.L.M., Dimitriadou, M.L., De Pauw, N., 2007. Decision tree models for prediction of macroinvertebrate taxa in the river Axios (Northern Greece). Aquatic Ecology 41, 399–411.

De Cooman, W., Peeters, B., Theuns, I., Vos, G., Lammens, S., Debbaudt, W., Timmermans, G., Meers, B., Van Erdeghem, M., Van Wauwe, P., Callebaut, R., Barrez, I., Van den Broeck, S., Emery, J., Van Volsem, S., Bursens, K., Van Hoof, K., D'Heygere, T., Soetaert,

H., Martens, K., Baten, I., Goris, M., Haustraete, K., Breine, J., Van Thuyne, G., Belpaire, C., Smis, A., 2007. Milieurapport Vlaanderen, MIRA Achtergronddocument 2007, Kwaliteit oppervlaktewater. [in Dutch] Vlaamse Milieumaatschappij, Aalst.

Dominguez-Granda, L., Lock, K., Goethals, P.L.M., 2011. Application of classification trees to determine biological and chemical indicators for river assessment: case-study in the Chaguana watershed (Ecuador). Journal of Hydroinformatics 14, 489–499.

European Council, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Communities L327, 22 December 2000, Brussels.

Gabriels, W., Lock, K., De Pauw, N., Goethals, P.L.M., 2010. Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnologica 40, 199–207.

Hoang, T.H., Lock, K., Mouton, A., Goethals, P.L.M., 2010. Application of decision trees and support vector machines to model the presence of macroinvertebrates in rivers in Vietnam. Ecological Informatics 5, 140–146.

Jahnig, S.C., Brabec, K., Buffagni, A., Erba, S., Lorenz, A.W., Ofenbock, T., Verdonschot, P.F.M., Hering, D., 2010. A comparative analysis of restoration measures and their effects on hydromorphology and benthic invertebrates in 26 central and southern European rivers. Journal of Applied Ecology 47, 671–680.

Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K., 2001. Improvements to Platt's SMO algorithm for SVM classifier design. Neural Computation 13, 637–649.

Lock, K., Goethals, P.L.M., 2011. Distribution and ecology of the mayflies (Ephemeroptera) of Flanders (Belgium). Annales de Limnologie - International Journal of Limnology 47, 59–165.

Meers, E., Tack, F.M.G., Tolpe, I., Michels, E., 2008. Application of a full-scale constructed wetland for tertiary treatment of piggery manure: monitoring results. Water, Air, and Soil Pollution 193, 15–24.

Mouton, A.M., Van der Most, H., Jeuken, A., Goethals, P.L.M., De Pauw, N., 2008. Evaluation of river basin restoration options by the application of the water framework directive explorer in the Zwalm River basin (Flanders, Belgium). River Research and Applications 23, 1–16.

Mouton, A.M., De Baets, B., Goethals, P.L.M., 2009. Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environmental Modelling and Software 24, 982–993.

Pino-Mejias, R., Cubiles-de-la-Vega, M.D., Anaya-Romero, M., Pascual-Acosta, A., Jordan-Lopez, A., Bellinfante-Crocci, N., 2010. Predicting the potential habitat of oaks with data mining models and the R system. Environmental Modelling and Software 25, 826–836.

Sahu, M., Gu, R.R., 2009. Modeling the effects of riparian buffer zone and contour strips on stream water quality. Ecological Engineering 35, 1167–1177.

Sponseller, R.A., Benfield, E.F., Valett, H.M., 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. Freshwater Biology 46, 1409–1424.

StatSoft Inc., 2004. STATISTICA (data analysis software system), version 7. Statsoft, Tulsa www.statsoft.com.

Strayer, D.L., Beighley, R.E., Thompson, L.C., Brooks, S., Nilsson, C., Pinay, G., Naiman, R.J., 2003. Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. Ecosystems 6, 407–423.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M., 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.

Tixier, G., Felten, V., Guérold, F., 2009. Life cycle strategies of *Baetis* species (Ephemeroptera, Baetidae) in acidified streams and implications for recovery. Fundamental Applied Limnology 174, 227–243.

Tran, C.P., Bode, R.W., Smith, A.J., Kleppel, G.S., 2010. Land-use proximity as a basis for assessing stream water quality in New York State (USA). Ecological Indicators 10, 727–733.

VMM, 2009a. Indicator report 2008. Van Steertegem M. (red.), Flanders Environment Report, Flemish Environment Agency, Aalst.

VMM, 2009b. Flanders environment outlook 2030. Van Steertegem M. (red.), Flanders Environment Report, Flemish Environment Agency, Aalst.

VMM, 2010. Jaarrapport water 2009. [in Dutch] Flemish Environment Agency, Aalst. http://www.vmm.be/pub/jaarrapport-water-2009.

Wasson, J.G., Villeneuve, B., Iital, A., Murray-Bligh, J., Dobiasova, M., Bacikova, S., Timm, H., Pella, H., Menging, N., Chandesris, A., 2010. Large-scale relationships between basin and riparian land cover and the ecological status of European rivers. Freshwater Biology 55, 1465–1482.

Witten, I.H., Frank, E., 2005. Data mining: practical machine learning tools and techniques. Morgan Kaufmann Publishers, San Francisco.